

Inpainting in Shape from Focus: Taking a Cue from Motion Parallax

Rajiv R. Sahay
sahayitm@gmail.com

A. N. Rajagopalan
raju@ee.iitm.ac.in

Image Processing and Computer Vision
Lab

Department of Electrical Engineering
Indian Institute of Technology Madras
Chennai 600 036, India

Abstract

Shape from focus (SFF) which uses a sequence of space-variantly defocused frames works under the constraint that there is ‘no magnification’ in the stack. In the presence of sensor damage and/or occlusions, there will be missing data in the observations and SFF cannot recover structure in those regions. In many applications, the capability of filling-in missing data is of critical importance. In this paper, we investigate the effect of motion parallax in SFF and demonstrate the interesting possibility of how it can be judiciously used to jointly inpaint image and depth profiles. When there is relative motion between the 3D specimen and the camera, by virtue of the inherent pixel motion in each of the frames, it is possible to obtain a focused image and depth map of the scene despite missing regions in the observations.

1 Introduction

Shape from focus (SFF) [24] is a passive ranging technique wherein a stack of images are captured using a real-aperture camera. The technique is based upon measuring the degree of focus locally in a stack of space-variantly blurred observations to arrive at the structure of the 3D scene. It makes the critical assumption of no pixel motion to be able to apply a window and compute the focus measure profile for every pixel through the stack. The relative motion between the scene and the optical system is limited to be small so as to avoid magnification effects among the frames. Researchers acknowledge the fact that magnification will cause errors in depth estimation in methods that use the blur cue and have devised methods to avoid it.

In the related area of depth from defocus (DFD), telecentricity on the image side has been proposed [33] to avoid any magnification effects. Object side telecentricity has been known in the literature [18] but in order to achieve this effect the aperture of telecentric lenses needs to be as large as the object to be viewed. This requires large, heavy and expensive lenses.

Existing works assume that there are no missing data in the frames. In real-world situations, the observations captured may be incomplete due to faulty imaging sensor or due to occlusions. Techniques such as SFF and DFD cannot recover the estimates of structure in regions where the image data is missing. We observe that, in such a scenario, motion parallax provides a vital cue for filling-in missing data. This is motivated by the fact that

missing/occluded portions of an object reappear in different observations when there are magnification effects. Here, we use the term magnification or scaling to loosely refer to structure-dependent pixel motion. In the proposed work, instead of shunning magnification, we exploit the pixel motion cue in the observations to expand the scope of SFF to inpaint the depth profile and the focused image of the underlying specimen despite missing information in the captured observations.

Several algorithms have been proposed for filling-up missing areas in digital images [16], [22], [2]. Image inpainting assumes that the area to be inpainted is known a priori and uses information in the neighbourhood to recover plausible details in the damaged regions. This technique has several applications which include restoration of scanned old photographs, astronomical images, removal of text/subtitles in images as well as in videos, and also for disocclusion/removal of objects in the frames. In [2], partial differential equations are used to propagate the boundaries in the direction of the isophotes. The algorithm works by populating both the gradient-direction (geometry) and the grayscale intensities (photometry) inside the inpainting region. Extending the techniques from 2D to 3D data, this idea has also been applied successfully for hole-filling during surface reconstruction [31]. The unification of structure and texture inpainting has been used by [1] where they split the input image into two different images, one capturing the basic image structure and the other capturing the texture information. The ideas of texture synthesis and image inpainting have also been combined together elegantly in [9] to fill large gaps. A similar formulation has been used for inpainting in video sequences as well [26, 27]. The work in [7] exploits geometric image models such as the bounded variation image model in a Bayesian framework. This approach is improved in [10] where the Mumford-Shah-Euler image model is proposed as the prior. Recently, in [14] it has been proposed to compute the 3D shape by controlling focus and aperture of a single camera especially for scenes that are highly geometrically complex and have fine texture. However, this method uses sophisticated SLR cameras and captures hundreds of high-resolution images to derive the structure. Its performance suffers for regions of low texture in the scene and the authors of [14] have resorted to inpainting using the technique in [2]. Importantly, in their work [14] observations are not affected by parallax as both the camera and the scene are stationary. Several works have been proposed recently in the DFD framework [11, 13, 17]. However, none of these works use frames with missing data/occlusions and affected by motion parallax. Occlusions have been handled for estimating depth information in [4, 12, 23] but a major difference between our approach and these works is that they consider only equifocal planes. The work in [15] does not address occlusion removal per se but only accounts for occlusion effects at defocused boundaries. We remark that in all the above works the observations are not affected by parallax and their formulations do not account for it.

In a recent paper [25], joint super-resolution of video and inpainting has been demonstrated. The inverse problem is solved using a total variation based regularization approach thereby tackling the inpainting problem also. In another recent work [32], joint depth completion and image inpainting has been attempted. The authors have shown that it is possible to handle occlusions in a stereo setup by inpainting both the depth and the texture. The works in [25, 32] are the closest in comparison to our approach. However, there are several differences. The work in [25] is about super-resolution and inpainting of image data but without any occlusions. Note that the authors of [25] do not reconstruct the depth map of the scene. The work in [32], is based on stereo and requires focused images. Our work clearly supersedes that of [25] and [32].

In the literature, image/video inpainting has primarily been attempted using focused im-

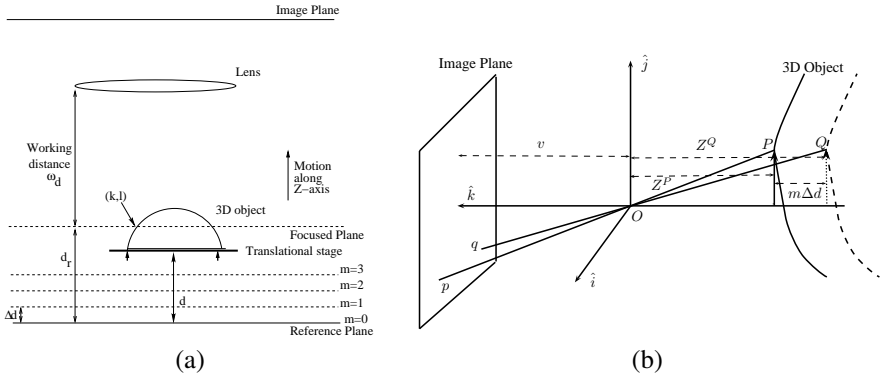


Figure 1: (a) Working principle of SFF. (b) Schematic showing mechanism of structure-dependent pixel motion in SFF.

ages without any magnification or parallax. The method proposed in this paper simultaneously reconstructs the inpainted depth map and the focused image from scaled and space-variantly blurred images. In this work, the formation of observations containing missing regions is depicted using an appropriate degradation model. A Bayesian approach is adopted in which the inpainted depth profile and the focused image are modeled as separate Gauss-Markov random fields (MRFs). Their maximum a posteriori (MAP) estimates are obtained by minimizing a suitably derived objective function.

In section 2, we motivate the need for inpainting in SFF. Pixel motion due to parallax effect is discussed in section 3. The degradation model for missing observations is explained in section 4. The proposed method for joint image and depth inpainting is given in section 5. Experimental results are described in section 6. Concluding remarks are presented in section 7.

2 Inpainting in SFF

The basic principle of SFF is depicted in Fig. 1 (a). A real-aperture camera images an object which is moved along the optical axis in finite fixed steps of Δd . Due to the finite depth-of-field of the camera and the 3D nature of the underlying specimen, the frames in the stack are space-variantly blurred. Initially, the object is placed on the ‘reference plane’ at a distance d_r from the ‘focused plane’ and moved upwards such that at every step a space-variantly blurred image is captured. As the point (k, l) approaches the ‘focused plane’ it gradually begins to get focused. When the distance between the reference plane and the translating stage is $d = m\Delta d = \bar{d}(k, l)$, the point (k, l) is on the ‘focused plane’ (at a distance of w_d from the lens plane) and is in perfect focus (neglecting diffraction effects and aberrations) since it satisfies the lens law. The fundamental idea behind the SFF technique is to estimate for every point the position of best focus, $\bar{d}(k, l)$. The variable $\bar{\mathbf{d}} = \bar{d}(k, l)$ for all the points (k, l) on the 3D specimen characterizes its shape. To detect the position of best focus for a point on the 3D object a focus measure operator is used to measure the quality of image focus in a local region. Specifically, the sum-modified Laplacian (SML) operator [24] is used to compute a focus measure profile for each pixel in the frame.

The estimate of the shape of the object is obtained by Gaussian interpolation. Note that in the above formulation it is assumed that the pixel locations do not change between frames.

This is crucial since the window used for computing the SML is centered around the same point in all the frames. In real-world situations, data in portions of the observations may be missing due to several factors including damage to the CCD sensor or occlusions. Since, the SFF analysis is local and depth estimates for a particular point are computed using data in the observations in the immediate neighbourhood of that pixel, this technique cannot recover structure in missing regions. Interestingly, we show that when there is magnification in the stack, structure-dependent pixel motion can be tapped to perform inpainting of both the depth profile as well as the focused image of the 3D specimen.

3 Motion parallax in the stack

When the relative motion between scene and camera is significant, there will be magnification (due to motion parallax) in the stack, such that the image features shift spatially from frame to frame. To explain the mechanism of this structure-dependent pixel motion, let us assume that the focused image and the shape of the 3D specimen are known.

Let the 3D specimen move relative to the camera along the axial direction as a sequence of frames is captured. Initially, we consider a pinhole camera and describe how scaled images are formed. As shown in Fig. 1 (b), we consider a specific point on the specimen which is moved relative to the pinhole camera. A point on the 3D object with world coordinates $P(X^P, Y^P, Z^P)$ moves to $Q(X^Q, Y^Q, Z^Q)$ along the Z-axis by a distance of $m\Delta d$ and away from the pinhole denoted by O . The distances of the points P and Q from the pinhole are Z^P and Z^Q , respectively. The point P is imaged at p on the image plane and has coordinates (x, y) . Let this image be the reference plane. When the 3D object is moved away from the pinhole by an amount $m\Delta d$, the point Q is imaged at q with coordinates (x', y') on the image plane. The corresponding image is the m^{th} frame.

Assuming that the size of the images is $M \times M$, according to basic perspective projection equations

$$x = \frac{vX^P}{Z^P}, \quad x' = \frac{vX^Q}{Z^Q} \quad \text{and} \quad y = \frac{vY^P}{Z^P}, \quad y' = \frac{vY^Q}{Z^Q} \quad (1)$$

It must be noted that the motion of the object relative to the pinhole is only along the Z-axis since the 3D specimen is translated away from or towards the camera along the optical axis. Hence, for the SFF scenario, we have $X^P = X^Q$, $Y^P = Y^Q$ and $Z^Q = Z^P + m\Delta d = w_d + \bar{d} + m\Delta d$, where v is the distance between the pinhole and the image plane, and $\frac{-M}{2} \leq x', y' \leq \frac{M}{2}$. Thus, it can be shown that

$$x' = \frac{x(w_d + \bar{d})}{(w_d + \bar{d}) + m\Delta d}, \quad y' = \frac{y(w_d + \bar{d})}{(w_d + \bar{d}) + m\Delta d} \quad (2)$$

Note that the pixel motion is a function of \bar{d} , the 3D structure of the scene. We will exploit this cue for inpainting.

4 Degradation model

Consider N frames, $\{y_m(i, j)\}$, $m = 0, 1, \dots, N-1$, each of size $M \times M$ from the stack. Assume that these are derived from a single focused image $\{x(i, j)\}$ of the 3D specimen. The

scaled and defocused frames can be related to the focused image by the degradation model

$$\mathbf{y}_m^{\text{vis}} = \mathbf{O}_m[\mathbf{H}_m(\bar{\mathbf{d}})\mathbf{W}_m(\bar{\mathbf{d}})\mathbf{x} + \mathbf{n}_m] \quad m = 0, \dots, N-1 \quad (3)$$

where $\mathbf{y}_m^{\text{vis}}$ is the lexicographically arranged vector of size $M^2 \times 1$ derived from the visible regions in the m^{th} defocused and scaled frame, \mathbf{W}_m is the matrix describing the motion of the pixels in the m^{th} frame, \mathbf{H}_m is the blurring matrix for the m^{th} frame, \mathbf{n}_m is the $M^2 \times 1$ Gaussian noise vector in the m^{th} observation and \mathbf{O}_m is the operator which removes the missing/damaged regions and crops out the visible portions of the observations. Note that for both the cases of sensor damage or occlusions (due to a static occluder), the cropping operator is unchanged for different frames of the stack, i.e., $\mathbf{O}_m = \mathbf{O}$. Hence, the spatial locations of the inpainting region remain the same in all the frames.

Note the interesting relationship between the shape of the object, pixel-motion and space-variant defocusing in the frames. The degree of space-variant defocus blur induced at each point in the image of a 3D scene is dependent upon the depth of the object from the lens plane. Also, the pixel motion is a function of the 3D structure of the object. In fact, the twin cues of defocus and motion parallax are intertwined with the 3D structure and must be judiciously exploited.

In the previous section, we have elucidated the motion parallax effect as modeled by $\mathbf{W}_m(\bar{\mathbf{d}})$ in (3). Next, we describe the space-variant blurring mechanism as specified by $\mathbf{H}_m(\bar{\mathbf{d}})$ in (3). In its most general form, the image formation process (as described by (3)) is quite difficult to comprehend and some simplifying assumptions must be made to gain a handle over the problem.

Towards this end, we assume a parametric model for the point spread function (PSF) of the camera. The PSF of a camera defines its response to a point light source. Due to diffraction and lens aberrations, the PSF is best described by a circularly symmetric 2D Gaussian function [28] with standard deviation $\sigma = \rho r_b$ where ρ is a camera constant and r_b is the blur circle radius. There exist several works that have validated this approximation [8], [5] and hence, we too are motivated to use this model. Since the blur parameter σ is a function of depth, the blurring induced is space-variant.

When the stage is moved upwards in steps of Δd (see Fig. 1), for the m^{th} frame we can express the blur parameter for a 3D point whose image pixel coordinates are (k, l) as

$$\sigma_m(k, l) = \rho R v \left(\frac{1}{w_d} - \frac{1}{w_d - m\Delta d + \bar{d}(k, l)} \right) \quad (4)$$

where $\frac{1}{w_d} = \frac{1}{f} - \frac{1}{v}$, f is the focal length, R is radius of aperture of the lens and v is the distance between the lens and the image plane.

Since the object is moving upwards as shown in Fig. 1, we have a negative sign for the term $m\Delta d$ in (4). Here, $\bar{d}(k, l)$ is the distance by which the stage must be moved from the reference plane to bring point (k, l) into focus.

If the separation between the stage and the reference plane is such that $d = m\Delta d = \bar{d}(k, l)$, then this point satisfies lens law and will appear in perfect focus. Note that under this condition, the blur parameter becomes zero, as expected (assuming the effects of optical aberrations to be negligible). We can establish a relationship between the blur parameter at a point in the reference frame with the blur parameter at the same point in the m^{th} frame of the stack as

$$\sigma_m(k, l) = \sigma_0(k, l) + \frac{\rho R v}{w_d + \bar{d}(k, l)} - \frac{\rho R v}{w_d + \bar{d}(k, l) - m\Delta d} \quad (5)$$

The product ρRv can be found using an appropriate calibration procedure and it remains constant during the entire image capturing process

5 Proposed framework for inpainting/disocclusion

In the previous section, we described the interaction of space-variant blurring and motion parallax in the frames when there is significant relative motion between the 3D specimen and the real-aperture camera. The problem that we attempt to solve is the inverse: Given a stack of scaled and defocused observations wherein there are regions of missing data due to possible occlusions in the scene or due to faulty camera sensor, how do we simultaneously inpaint the depth profile and the focused image of the object?

We propose to inpaint both the depth profile as well as the focused image of the object within a unified framework. Note that for the problem considered in this work, the spatial locations of the missing regions do not change from frame to frame in the stack. It is probable that whatever region of the 3D specimen was occluded/missing in one observation of the stack will come into view in another frame since there is magnification in the stack.

Simultaneous reconstruction of the depth profile $\bar{\mathbf{d}}$ and the focused image \mathbf{x} is an ill-posed inverse problem and hence, the solution has to be regularized using a priori constraints. Real-world objects usually have depth profiles which are locally smooth. The same argument holds good for the focused image also. Markov random fields (MRFs) have the capability to model spatial dependencies so as to incorporate prior information [21]. We model the structure and the focused image of the 3D specimen using separate Gauss-Markov random fields (GMRF) with a first-order neighbourhood. The Hammersley-Clifford theorem [3] provides the all-important equivalence between Gibbs random field and MRF, enabling the specification of the prior joint pdf for the depth map and the focused image. For details on MRF, see [21].

We seek to solve for the MAP estimate of $\bar{\mathbf{d}}$ and \mathbf{x} . Let us consider a set of p frames chosen from the stack of N observations. Assuming the noise process \mathbf{n}_m 's to be independent, the MAP estimates of $\bar{\mathbf{d}}$ and \mathbf{x} can be obtained by minimizing the posterior energy function

$$U^p(\bar{\mathbf{d}}, \mathbf{x}) = \sum_{m \in O} \frac{\|\mathbf{y}_m^{\text{vis}} - \mathbf{O}_m[\mathbf{H}_m(\bar{\mathbf{d}})\mathbf{W}_m(\bar{\mathbf{d}})\mathbf{x}]\|^2}{2\sigma_\eta^2} + \lambda_{\bar{\mathbf{d}}} \sum_{c \in C_{\bar{\mathbf{d}}}} V_c^{\bar{\mathbf{d}}}(\bar{\mathbf{d}}) + \lambda_{\mathbf{x}} \sum_{c \in C_{\mathbf{x}}} V_c^{\mathbf{x}}(\mathbf{x}) \quad (6)$$

where $O = \{u_1, u_2, \dots, u_p\}$, u_i is the frame number and σ_η^2 is the variance of the Gaussian noise. The clique potential function for $\bar{\mathbf{d}}$ is

$$\sum_{c \in C_{\bar{\mathbf{d}}}} V_c^{\bar{\mathbf{d}}}(\bar{\mathbf{d}}) = \sum_{i=1}^M \sum_{j=1}^M [(\bar{d}(i, j) - \bar{d}(i, j-1))^2 + (\bar{d}(i, j+1) - \bar{d}(i, j))^2 + (\bar{d}(i+1, j) - \bar{d}(i, j))^2 + (\bar{d}(i, j) - \bar{d}(i-1, j))^2] \quad (7)$$

where c is a clique, C is the set of all cliques and $V_c(\cdot)$ is the potential associated with clique c . We choose $V_c^{\mathbf{x}}(\mathbf{x})$ to be of the same form to $V_c^{\bar{\mathbf{d}}}(\bar{\mathbf{d}})$.

Recently, graph cuts have been proposed as a fast optimization technique for several computer vision applications. However, they are limited to minimization of submodular energy functions such as used in areas like image segmentation, stereo matching etc. [6] [20]. In applications that involve blur, the cost functions in the MAP-MRF framework turn

out to be non-submodular [29],[30]. For such energy functions, graph cuts have not been shown to exceed the performance of simulated annealing (SA) [30], [19]. The usefulness of the Quadratic Pseudo Boolean Optimization (QPBO) algorithm depends upon how many nodes in the graph are labeled [30] (page 3, (section 2.2)). Also, in section 3 of [30], it is asserted that the roof duality works well in cases where the number of non-submodular terms is small. However, in more difficult cases the roof duality technique leaves many nodes unassigned. We refer to Table 1, on page 7 of [30]. The comparison results for image deconvolution (only ‘space-invariant’ blurring with number of gray levels limited to 32) are presented in the last two rows (3×3 and 5×5 kernels) of this table. Note that the number of unassigned labels for the QPBO and the ‘probing’ QPBO (QPBO) methods are a whopping 80% for the 5×5 sized kernel. Note that using SA, the energy at convergence is zero and all the nodes are labeled. Interestingly, SA is shown to outperform all the methods including QPBO and QPBOP (pointed out in [30] in the ‘Image Deconvolution’ sub-section on Page 8). Hence, we have used the SA algorithm to minimize $U^p(\bar{\mathbf{d}}, \mathbf{x})$. Parameters $\lambda_{\bar{\mathbf{d}}}$ and $\lambda_{\mathbf{x}}$ must be tuned to obtain a good estimate of both $\bar{\mathbf{d}}$ and \mathbf{x} .

6 Experimental results

We captured sequences of scaled and space-variantly defocused images using an off-the-shelf Olympus C5050Z digital camera which was operated in the ‘super-macro’ mode. To simulate the effect of missing data due to possible damage to the CCD sensor, we randomly scratched the captured frames. Note that during the inpainting process we assume the knowledge of the locations of the regions which have to be filled in. This is a standard assumption in the literature and we exploit it in our framework. For all our experiments, we feed the entire stack as input to the method in [24], to obtain the initial estimate of the inpainted depth map. For the inpainted focused image, we choose an arbitrary initial estimate as a cropped portion of the Lena image. In all the experiments, frames are captured by moving the object along the optical axis in finite steps of $\Delta d = 1$ mm.

In the first experiment, we chose a real-world specimen, a small wooden piece on which a face is carved. The captured images are space-variantly blurred and scaled. Note the locations indicated by white arrows in Fig. 2 (a). At these locations, in Fig. 2 (b) we can make the interesting observation that the position of the damaged regions is fixed but different image features are covered by the black pixels in this image. This is the cue which we exploit to fill-in the missing parts in both the focused image and the depth profile. The inpainted image using the proposed algorithm is shown in Fig. 3 (a). The values of the parameters chosen in the optimization procedure were $\lambda_{\bar{\mathbf{d}}} = 1 \times 10^8$ and $\lambda_{\mathbf{x}} = 0.05$, respectively. The reconstructed depth map is shown in Fig. 3 (b). We can see that the proposed method is able to reconstruct the depth information even in regions with missing observations. It is notable that various features of the face like the eyes, the eyebrows, the nose and the straight edge below it are faithfully reconstructed both in the inpainted image (Fig. 3 (a)) and the inpainted depth map (Fig. 3 (b)), respectively.

In the next experiment, we chose another real-world specimen, a small clay model of a bunny. Two frames of a portion of the head of the bunny are shown in Fig. 4 (a) and (b), respectively. To simulate the regions where there is sensor damage, we have marked them as black in the images. The proposed algorithm was run with four frames from the stack as the input. The inpainted focused image of the specimen is shown in Fig. 5 (a). Several details that were missing are recovered. The algorithm is able to propagate the information

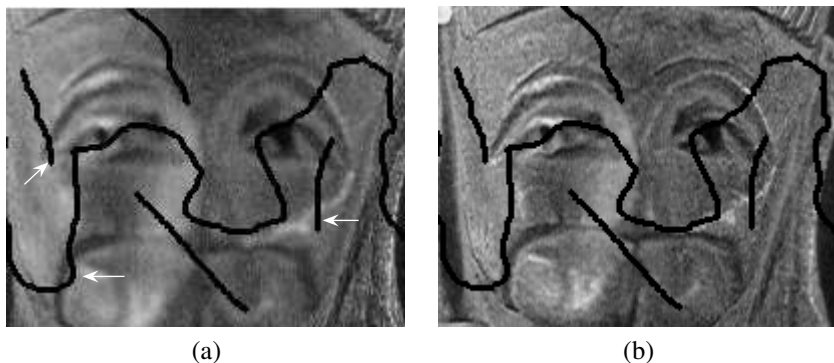


Figure 2: Wooden specimen depicting a face. (a, b) Two frames chosen from the scaled and defocused stack with missing data, simulating the effects of a camera having a damaged sensor.

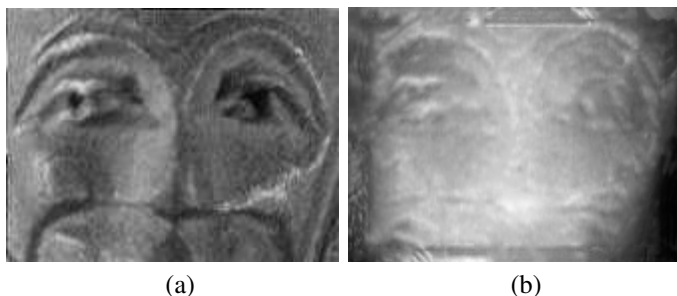


Figure 3: (a) Inpainted focused image. (b) Grayscale image corresponding to the inpainted depth map.

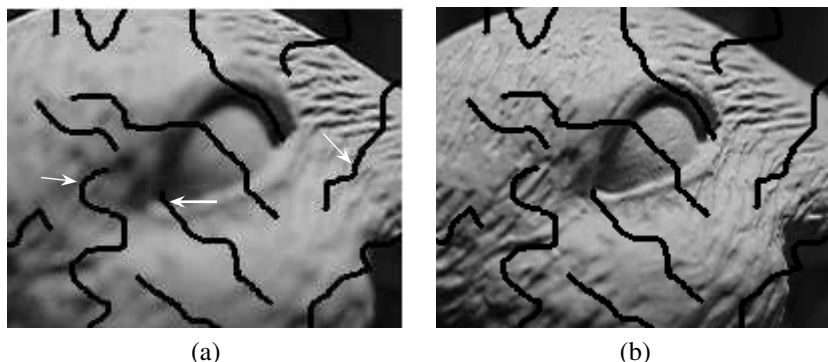


Figure 4: Clay model of a bunny. (a, b) Second and fourth frames showing the effects of magnification and space-variant defocusing and regions of missing data.

in the smooth regions into the gaps as well as preserve the edges. The grayscale image of the reconstructed depth profile is depicted in Fig. 5 (b). Observe that the contours of the eyelids have been recovered well.

To demonstrate the capability of our method in performing disocclusion, we present results of yet another experiment. Here, a small pin was kept across the field-of-view of the

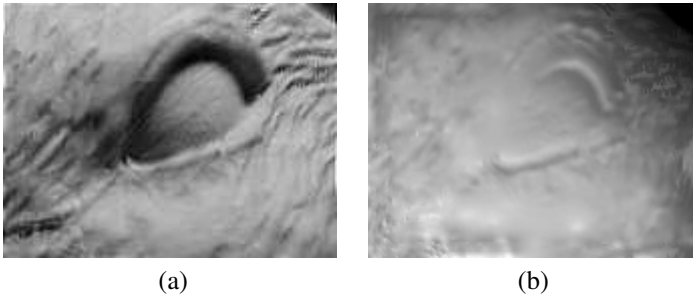


Figure 5: (a) Inpainted focused image of the bunny specimen. (b) Grayscale image of the inpainted depth profile obtained using the proposed method.

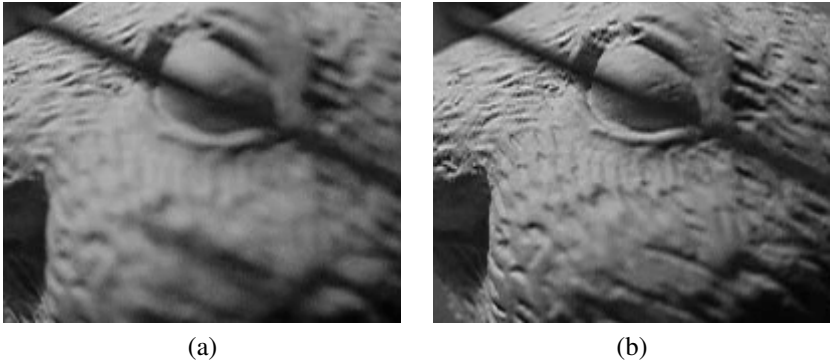


Figure 6: A clay model of a bunny occluded by a straight object. (a, b) Two frames chosen from the stack.

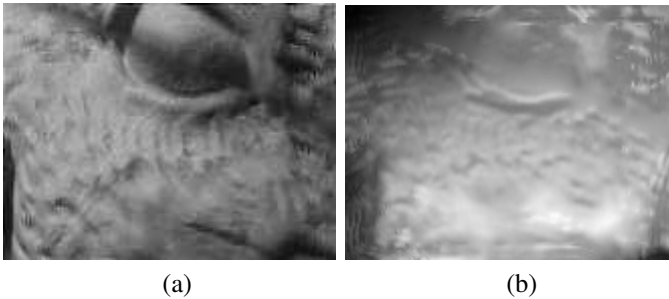


Figure 7: (a) Inpainted focused image using our method demonstrating disocclusion. (b) Grayscale image of the shape profile.

camera such that it occludes the specimen which is chosen as a clay model of a bunny. We have provided as supplementary material, a video captured as the specimen is moved in front of the static camera and occluder. To demonstrate the effects of occlusion, scaling and space-variant defocusing, two frames from the stack are shown in Fig. 6 (a) and (b), respectively. Note that the position of the occluded regions is fixed in all the observations and is assumed to be known. The proposed method yields the inpainted focused image shown in Fig. 7 (a). The features of the eye portion at the locations of the occluder have been recovered well. Note that the edges that were occluded have been recovered. Particularly, the edges at the

junction of the eyelids and the eyeball are clearly visible. Also, the texture on the surface of the specimen has been deblurred well and can be discerned easily. The estimated depth profile is shown in Fig. 7 (b). Even in occluded regions, the algorithm is able to recover the depth information. We used four frames with good relative blur among them from the stack to reconstruct the completed focused image and the depth map in all the experiments.

7 Conclusions

In this work, we demonstrated the interesting possibility of using motion parallax as a cue for inpainting in the SFF scenario. When a real-world camera is used to capture a stack, the frames are subject to magnification effects. In addition, there can be data loss in certain regions due to sensor damage and/or occlusions. Despite missing regions in the observations, we have shown that it is possible to obtain the focused image and reconstruct the structure of the underlying specimen using only a few frames from the stack. In effect, we are able to integrate inpainting, deblurring and structure recovery within a single unified framework.

References

- [1] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and image inpainting. *IEEE Trans. Image Processing*, 12(8):882–889, 2003.
- [2] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. Computer Graphics, SIGGRAPH 2000*, pages 417–424, 2000.
- [3] J. Besag. Spatial interaction and the statistical analysis of lattice system. *J. Royal Statistical Soc.*, pages 192–236, 1974.
- [4] S. S. Bhasin and S. Chaudhuri. Depth from defocus in presence of partial self occlusion. In *Proc. ICCV*, pages 488–493, 2001.
- [5] M. Born and E. Wolf. *Principles of optics*. Pergamon, London, 1965.
- [6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [7] T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM Journal of Applied Mathematics*, 62:1019–1043, 2001.
- [8] S. Chaudhuri and A. N. Rajagopalan. *Depth from defocus: A real aperture imaging approach*. Springer-Verlag, New York, 1999.
- [9] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Processing*, 13:1200–1212, 2004.
- [10] S. Esedoglu and J. Shen. Digital inpainting based on the mumford-shah-euler image model. *European Journal of Applied Mathematics*, 13:353–370, 2002.
- [11] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *IEEE Trans. on Pattern Anal. Mach. Intell.*, 27:406–416, 2005.

- [12] P. Favaro and S. Soatto. Seeing beyond occlusions (and other marvels of a finite lens aperture). In *Proc. CVPR*, pages 579–586, 2003.
- [13] P. Favaro, S. Soatto, M. Burger M., and S. J. Osher. Shape from defocus via diffusion. *IEEE Trans. on Pattern Anal. Mach. Intell.*, pages 518–531, 2008.
- [14] S. W. Hasinoff and K. N. Kutulakos. Confocal stereo. *International Journal of Computer Vision*, 81:82–104, 2009.
- [15] S. W. Hasinoff and K. N. Kutulakos. A layer-based restoration framework for variable-aperture photography. In *Proc. ICCV*, 2007.
- [16] A. Hirani and T. Totsuka. Combining frequency and spatial domain information for fast interactive image noise removal. In *Proc. Computer Graphics, SIGGRAPH 1996*, pages 269–276, 1996.
- [17] H. Jin and P. Favaro. A variational approach to shape from defocus. In *Proc. ECCV*, pages 18–30, 2002.
- [18] R. Kingslake. *Optical System Design*. Academic Press, New York, 1983.
- [19] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts—A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(7):1274–1279, 2007.
- [20] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.
- [21] S. Z. Li. *Markov random field modeling in computer vision*. Springer-Verlag, Tokyo, 1995.
- [22] S. Masnou and J. Morel. Level-lines based disocclusion. In *Proc. IEEE Intl. Conf. on Image Processing*, pages 259–263, 1998.
- [23] M. Mcguire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand. Defocus video matting. In *Proc. ACM SIGGRAPH*, pages 567–576, 2005.
- [24] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994.
- [25] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang. A total variation regularization based super-resolution reconstruction algorithm for digital video. *EURASIP Journal on Advances in Signal Processing*, 2007(74585):16, 2007.
- [26] K. A. Patwardhan, G. Sapiro, and M. Bertalmio. Video inpainting of occluding and occluded objects. In *IEEE Intl. Conf. Image Processing*, pages 69–72, 2005.
- [27] K. A. Patwardhan, G. Sapiro, and M. Bertalmio. Video inpainting under constrained camera motion. *IEEE Trans. Image Processing*, 16(2):545–553, 2007.
- [28] A. P. Pentland. A new sense for depth of field. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(4):523–531, 1987.

- [29] A. Raj and R. Zabih. A graph cut algorithm for generalized image deconvolution. In *Proc. IEEE Intl. Conf. on Computer Vision*, pages 1048–1054, 2005.
- [30] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
- [31] J. Verdera, V. Caselles, M. Bertalmío, and G. Sapiro. Inpainting surface holes. In *Proc. IEEE Intl. Conf. on Image Processing*, pages 903–906, 2003.
- [32] L. Wang, H. Jin, R. Yang, and M. Gong. Stereoscopic inpainting: joint color and depth completion from stereo images. In *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
- [33] M. Watanabe and S.K. Nayar. Telecentric optics for focus analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(12):1360–1365, 1997.